Spatio-Temporal Data Fusion for Remote Sensing Applications

Amy Braverman[1] Hai Nguyen[1] Noel Cressie[2]
Matthias Katzfuss[2] Ed Olsen[1] Anna Michalak[3]

[1]Jet Propulsion Laboratory,
California Institute of Technology

[2]Department of Statistics,
The Ohio State University

[3]Department of Global Ecology, Carnegie Institution,
Stanford University

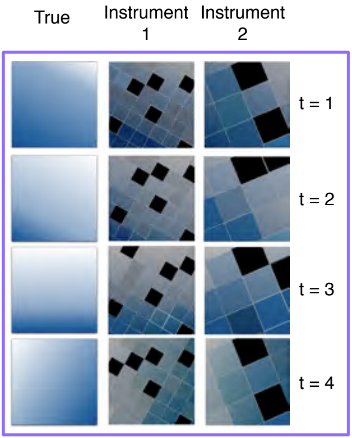
June 22, 2011

Outline

- Introduction
- Inference from spatial data.
- Inference from a single remote sensing data set.
- Inference from multiple remote sensing data sets: Spatial-Statistical Data Fusion (SSDF)
- Spatio-Temporal Data Fusion (STDF)
- STDF for two processes
- Application: infer CO2 in the lower-atmosphere from ACOS/GOSAT and AIRS.

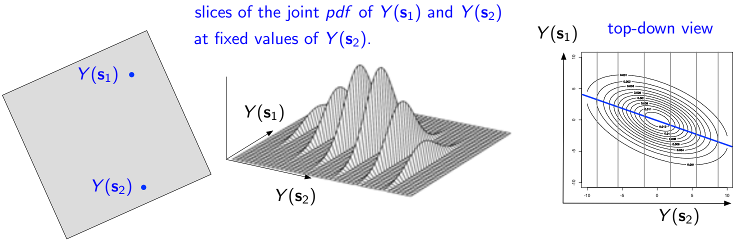
A family of inference problems:

Exploit spatial correlations	Exploit spatial and temporal correlations
Infer the true field (single quantity) from one remote sensing image of it at a single time point. (Fixed Rank kriging)	Infer the true field (single quantity) from one remote sensing image of it at multiple time points. (Fixed Rank filtering)
Infer the true field from *two* different remote sensing images of it at a single time. (Single process, multiple source spatial data fusion)	Infer the true field from *two* different remote sensing images of it at multiple time points. (Single process, multiple source spatio-temporal data fusion)
Infer true values of *two* fields from *two* different remote sensing images at a single time. (Muliple process, multiple source spatial data fusion)	Infer true values of *two* fields from *two* different remote sensing images at multiple time points. (Muliple process, multiple source spatio-temporal data fusion)



Inference from spatial data

- Let $\mathbf{s}_1$ and $\mathbf{s}_2$ be the (lat,lon) pairs of two point locations.
- Let $Y(\cdot)$ be a random variable representing the value of a quantity of interest at the location of its argument.

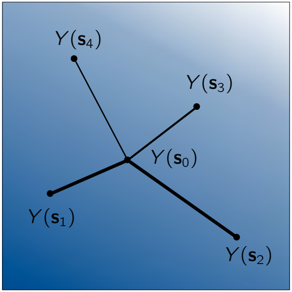


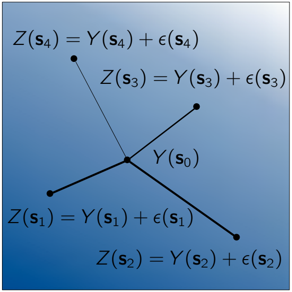
$E[Y(\mathbf{s}_1)|Y(\mathbf{s}_2)]$ = projection of the slice means onto the floor is a line (linear regression).

$E[\cdot]$ = expected value. $[Y(\mathbf{s}_1)|Y(\mathbf{s}_2)]$ = conditional distribution of $Y(\mathbf{s}_1)$ given $Y(\mathbf{s}_2)$.

Inference from spatial data

Kriging:





The best linear unbiased estimator of $Y(\mathbf{s}_0)$ is

$\hat{Y}(\mathbf{s}_0) = \mathbf{a}_{\mathbf{s}_0}{}'\mathbf{Z}$,

$\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), Z(\mathbf{s}_3), Z(\mathbf{s}_4))'$.

$\mathbf{a}$ minimizes $E(Y(\mathbf{s}_0) - \mathbf{a}'\mathbf{Z})^2$

subject to $E(\hat{Y}(\mathbf{s}_0)) = E(Y(\mathbf{s}_0))$.

$\mathbf{a} = \mathbf{c}(\mathbf{s}_0)\Sigma^{-1}$ *This could be a computational problem!*

$\mathbf{c}(\mathbf{s}_0) \equiv (C(\mathbf{s}_1, \mathbf{s}_0), \ldots, C(\mathbf{s}_4, \mathbf{s}_0))'$.

$Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = C(\mathbf{s}_i, \mathbf{s}_j)$ $\quad Cov(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = C(\mathbf{s}_i, \mathbf{s}_j)$, ($\epsilon$'s independent)

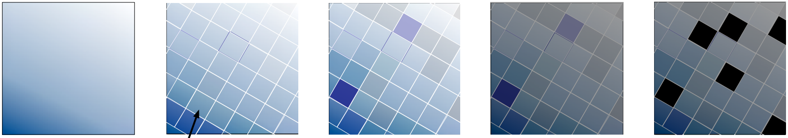
$\Sigma = [C(\mathbf{s}_i, \mathbf{s}_j)]$ $\quad Cov(Z(\mathbf{s}_i), Y(\mathbf{s}_0)) = C(\mathbf{s}_i, \mathbf{s}_0)$

Thickness of lines connecting locations indicates strength of spatial correlation.

Inference from remote sensing data

How to infer (A) given only (E)?

(A)	(B)	(C)	(D)	(E)
True	Discretized	With variance	With bias	With missing

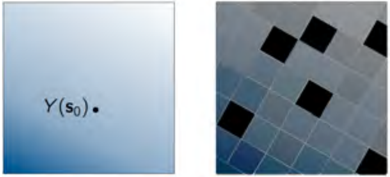


$Y(\mathbf{s})$ $\quad B_m$ $\quad \epsilon(B_m) \sim N(c, \sigma_\epsilon^2)$ $\quad N =$ number non-missing pixels

$$Y(B_m) = \frac{1}{|B_m|} \int_{\mathbf{s} \in B_m} Y(\mathbf{s})\, d\mathbf{s} \qquad Z(B_m) = Y(B_m) + \epsilon(B_m) \qquad \mathbf{Z} = (Z(B_1), \ldots, Z(B_N))'$$

- Remote sensing data are spatial aggregates over footprints.
- Inference desired at the point-level (on a fine grid of points, usually).

Change of Support and Fixed Rank Kriging



$\mathbf{Z} = (Z(B_1), \ldots, Z(B_N))'$

- *It is still true that the best linear unbiased estimator of $Y(\mathbf{s}_0)$ is*

$$\hat{Y}(\mathbf{s}_0) = \mathbf{a}'_{\mathbf{s}_0}\mathbf{Z}$$

even though the elements of $\mathbf{Z}$ are block values.

- $\mathbf{a}_{s_0}$ now involves covariances between points and blocks, and between blocks and blocks.
- Requires inversion of $N \times N$ covariance matrix $\mathbf{\Sigma}$, N may be very large (number of pixels).
- Usual assumptions of isotropy and stationarity are hard to justify.
- Alternative model that overcomes these problems is provided by Fixed Rank Kriging (FRK) and the Spatial Random Effects (SRE) model (Cressie and Johannesson, 2008).

Spatial-Statistical Data Fusion (SSDF)

- Two instruments: optimal estimator of $Y(\mathbf{s}_0)$ is $\hat{Y}(\mathbf{s}_0) = \mathbf{a}'_{1\mathbf{s}_0}\mathbf{Z}_1 + \mathbf{a}'_{2\mathbf{s}_0}\mathbf{Z}_2$.
- Requires estimating cross-covariances between blocks in different data sets, and estimating and correcting for different measurement error biases and variances.
- All that can be incorporated into the constrained minimization of $E\Big(Y(\mathbf{s}_0) - \hat{Y}(\mathbf{s}_0)\Big)^2$. See Nguyen (2009) for details.
- Conceptually, SSDF can be thought of as FRK on a combined dataset, $\mathbf{Z} = (\mathbf{Z}_1{}', \mathbf{Z}_2{}')'$, with change of support and bias correction. (Nguyen, Cressie, and Braverman, 2010).

The Spatial Random Effects Model

- Our interest is not really in the fusion coefficients (the **a**'s). It's in $\hat{Y}(\mathbf{s})$ and its standard error.

- An alternative formulation:

$$\underbrace{Y(\mathbf{s}) = \mu(\mathbf{s})}_{\text{Trend}} + \underbrace{\nu(\mathbf{s})}_{\text{Spatial covariance}} + \underbrace{\xi(\mathbf{s})}_{\text{Fine-scale variation}}.$$

$$\hat{Y}(\mathbf{s}) = \mu(\mathbf{s}) + \hat{\nu}(\mathbf{s}) + \hat{\xi}(\mathbf{s}).$$

Estimate these from *footprint-level* data from <u>*both*</u> sources.

- In particular, express $\nu(\mathbf{s})$ as a linear combination of elements of a hidden structure variable, $\boldsymbol{\eta}$:

$$\hat{\nu}(\mathbf{s}) = \mathbf{S}(\mathbf{s})'\hat{\boldsymbol{\eta}},$$

$\mathbf{S}(\mathbf{s})$ is a known weight vector for location $\mathbf{s}$, and $\boldsymbol{\eta}$ is low-dimensional. (Note: $Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = Cov(\nu(\mathbf{s}_i), \nu(\mathbf{s}_j))$.)

The Spatio-Temporal Random Effects Model

- Now we introduce time by letting $\boldsymbol{\eta}$ evolve according to an order 1 autoregressive process:

$$\boldsymbol{\eta}_{t+1} = \mathbf{H}_{t+1}\boldsymbol{\eta}_t + \boldsymbol{\zeta}_{t+1},$$

  where $\mathbf{H}$ is a non-stochastic “propagator” matrix, and ζ is an innovation vector that is independent of $\boldsymbol{\eta}$.

- At each time step we produce the “forecast” using the equation above, and use an empirical Bayesian formalism to update it after seeing the footprint-level data from both sources.

- We use a Kalman Filter on $\boldsymbol{\eta}$ to update its estimate as data for additional time points are acquired. (Technical point: the estimate of $\xi(\mathbf{s})$ also needs to be filtered since $\xi(\mathbf{s})$ is jointly distributed with the estimate of $\boldsymbol{\eta}$.)

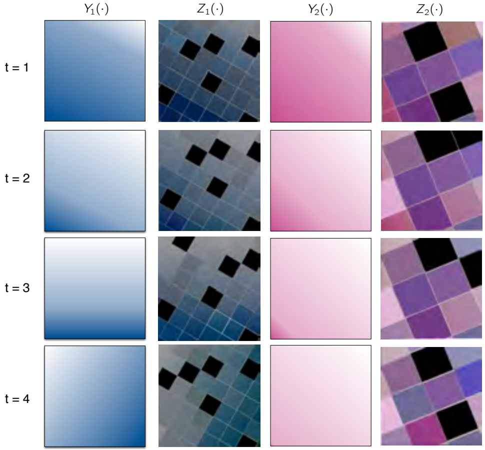
- $\hat{Y}(\mathbf{s}, t) = \mu_t(\mathbf{s}) + \mathbf{S}(\mathbf{s})'\hat{\boldsymbol{\eta}}_{t|t} + \hat{\xi}_{t|t}(\mathbf{s})$, “$t|t$” indicates using all data up to and including time t. This is called Fixed Rank Filtering (FRF; Cressie, Shi, and Kang, 2010).

Spatio-Temporal Data Fusion (STDF)

- Conceptually, STDF can be thought of as FRF on a combined dataset, $\mathbf{Z} = (\mathbf{Z}_1{}', \mathbf{Z}_2{}')'$, with change of support and bias correction.

- Coming soon: STDF based on Fixed Rank Smoothing (FRS; Katzfuss and Cressie, 2011) uses all the data during the period, not just data through time t. Allows estimates to be made for shorter time increments.

STDF for Vector Processes



Suppose the object of inference is a pair:

$$\mathbf{Y}(\mathbf{s}_0) = (Y_1(\mathbf{s}_0), Y_2(\mathbf{s}_0))'.$$

Everything generalizes but now there are inter-variable as well as inter- and intra-dataset covariances.

(N.B.: With more than two vector components, mathematics and computation become much more complex and intensive.)

Example: ACOS/GOSAT and AIRS CO2

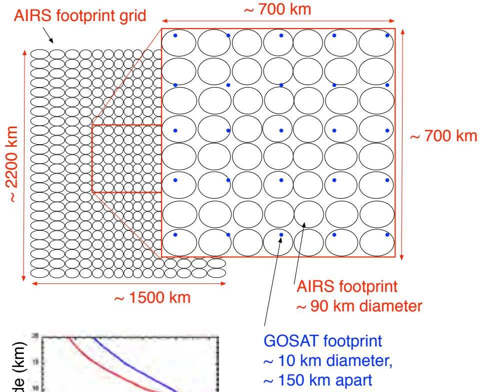
- $Y_1(\mathbf{s}_0, t)$ = total column CO2 volume mixing ratio (VMR).
- $Y_2(\mathbf{s}_0, t)$ = mid-troposphere VMR.
- Estimate:

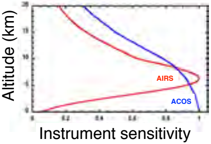
$$X = f_1 Y_1(\mathbf{s}_0, t) - f_2 Y_2(\mathbf{s}_0, t),$$

$$f_1 = (1000 - 300)/(1000 - 500),$$

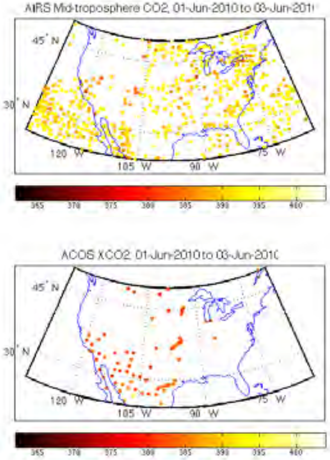
$$f_2 = (500 - 300)/(1000 - 500)$$

(f_1 and f_2 adjust for volume differences in pressure units.)

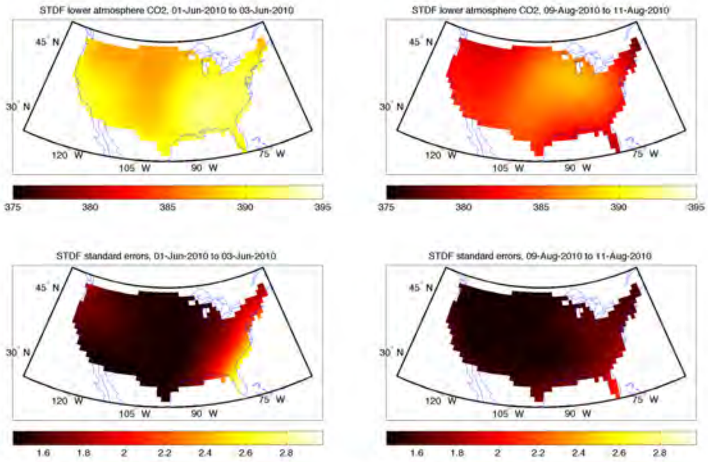




Example: ACOS/GOSAT and AIRS CO2



Example: ACOS/GOSAT and AIRS CO2



Example: GOSAT and AIRS CO2

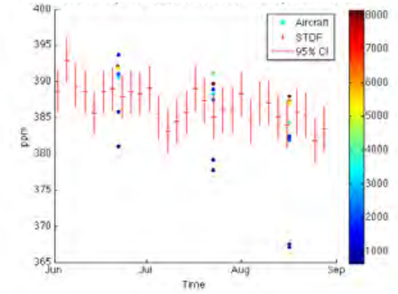
- 500-750 AIRS retrievals over the continental US every three days during summer 2010. Estimated measurement bias is zero, estimated measurement error standard deviation is 1.87 ppm.
- 90 ACOS retrievals over the continental US every three days during summer 2010. Estimated bias is -8 ppm, estimated measurement error standard deviation is 5 ppm.
- Estimation grid is $1^\circ \times 1^\circ$.
- These results are based on the FRS version of STDF, using the EM algorithm for estimation. (Results in the ESTF paper are based on FRF with binned method-of-moments estimation.)
- Computation time: about 24 minutes for this three-month analysis on 2.8 GHz MacBook Pro.

Validation (a start)

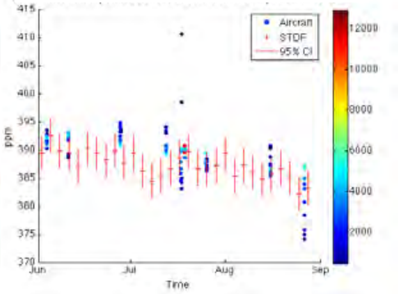
Comparison of STDF estimates with NOAA aircraft flights.

NOAA data courtesy of Colm Sweeney, ESRL

Beaver Crossing, NE



Lamont, OK



(Colorbars show aircraft sampling altitude.)

Conclusions

- FRS + EM = shorter time increments (3 days); better to capture dynamics.
- Just starting “validation” now. First results are not discouraging, but there are indications that our estimates are too low (especially outside of JJA).
- Need to incorporate instrument sensitivities into the formula for X.
- Method is fast: suitable for very large remote sensing data sets thanks to the STRE model.
- The ACOS/GOSAT and AIRS case is interesting because offers the possibility of combining two instruments’ data to derive an estimate of something neither one observes directly: CO2 concentrations in the lower atmosphere. This may be related to CO2 flux from the surface.

The end

Questions, comments?

Contact Amy.Braverman@jpl.nasa.gov.

Support for this work is provided by NASA's Earth Science Technology Office through its Advanced Information Systems Technology Program.

References

N. Cressie and G. Johannesson (2008). Fixed-Rank Kriging for Very Large Spatial Data Sets, Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 70, no. 1, pp 209-226.

H. Nguyen (2009). Spatial Statistical Data Fusion for Remote Sensing Applications, Ph.D. Dissertation, Department of Statistics, UCLA. Available at http://theses.stat.ucla.edu/abstract.php?dissertation=104

N. Cressie, T. Shi, and E. Kang, (2010) Fixed Rank Filtering for Spatio-Temporal Data, Journal of Computational and Graphical Statistics, vol. 19, no. 3, pp 724-745.

H. Nguyen, N. Cressie, and A. Braverman (2010). Spatial-statistical Data Fusion for Remote Sensing Applications, Journal of the American Statistical Association, submitted. (Also available as Technical Report No. 849, Department of Statistics, The Ohio State University, November, 2010.)

M. Katzfuss and N. Cressie (2011). Spatio-Temporal Smoothing and EM Estimation for Massive Remote-Sensing Data Sets, Journal of Time Series Analysis, Vol. 32, Issue 4, pp. 430-446.